

# THE FACTOR ANALYSIS AND THE SEARCH FOR AN OBJECTIVE MEANING OF FACTORS AS A FUNCTION OF MEANINGS OF (NAMES) FEATURES

V.D. Mazurov, vldmazurov@gmail.com

Ural Federal University named after the First President of Russia Boris Yeltsin, Ekaterinburg, Russian Federation

A method of the factor analysis, including factors naming is considered. If  $x$  – features, and  $f$  – factors, then we look for dependencies  $f(x)$ , not  $x(f)$ . In the beginning of the article there is a detailed historical overview of the origins and evolution of the theory of the factor analysis. Then methods of committee solutions of problems of pattern recognition are studying, including the discriminant analysis, taxonomy and informative subsystems of signs assessment. The connection of these methods with artificial neural networks and with the factor analysis that allows finding deep interconnections in observations table is analyzed.

A stepwise algorithm to find out the name of a factor is performed according to the names of features that are included in the corresponding taxon. Along with this the factor analysis is applied to the observations table of object\feature.

In conclusion, the full reference list is provided on the subject of a research.

*Keywords: factors, names, features, linguistics, statistics, algebra.*

Our approach to the theory of the factor analysis is closely connected with the theory of committee decisions, so I have decided to pay attention to the origins of this theory.

A factor – a latent source of dynamics of interdependent features of objects and phenomena.

The factor analysis is a method of multidimensional mathematical statistics. It is used in order to search for statistically associated features and for their assessment, in order to emphasize key hidden factors, functions of which are presented by features.

The founder of a factor analysis – English explorer, sir Francis Galton (1822–1911), geographer, anthropologist, psychologist, founder of differential psychology and psychometrics, statistician. In the 1850s he developed root ideas of a factor analysis implementing them into psychological problems of individual differences. The goal is to create a mathematical model of individual differences.

Then in 1901 English mathematician, statistician and biologist, Karl Pearson (1857–1936), founded mathematical statistics and biometrics; and suggested an idea of principal axis method.

English psychologist and mathematician Charles Edward Spearman (1863–1945) studied a two-factor model of human intelligence and distinguished a general factor.

The traditional procedure of finding the factors is performed through equations of the dependency of features from factors [1–3]. I have suggested a complex method – search for dependencies of factors from features.

The traditional approach includes factors taxonomy [4]. Along with this one relies on the statistics of learning material. In my approach we deal with algebraic procedures without mathematical statistics.

Mathematical models and methods of solutions committee dealing with pattern recognition problems are considered, including discriminant analysis, taxonomy and assessment of information capacity of signs subsystems [5]. Among the committee structures there is the committee of the majority which is the main one. This is one of the models of the experts' council. The main objective is to find a decision rule of pattern recognition.

The objective consists in the following. We need to find a committee of discriminant functions for the case of sets A and B. The discriminant function  $f$ , if it exists, satisfies the system of inequalities (\*):

$$F(a) > 0 \text{ for all } a \text{ from the set } A,$$

$$f(b) < 0 \text{ for all } b \text{ from the set } B,$$

$$f \text{ can be found in the functional class } F.$$

However, this system can often be inconsistent, and then, instead of one function we create the committee of  $C$  functions.

This is such a finite sequence

$$C = [f_1, \dots, f_q],$$

that more than half of the functions from the set  $C$  satisfy each inequality of the system (\*). Some of the functions of the set can be repeated.

The connection of these methods with the factor analysis is studied in this article which allows finding deep connections in the observations table, as well as with artificial neural networks. In contrast to traditional methods based on mathematical statistics (which require large amounts of observations and require search of dependencies of the features from the factors), we suggest an algebraic approach based on the committees method.

Our work is supported by RAS academician V.I. Berdyshev and by RAS academician Yu.I. Zhuravlev who is in charge of the entire area of algebraic models and methods of recognition and their mathematical and practical justification in the Russian Federation.

However, it is necessary to start with the initiative of the outstanding mathematicians S.B. Stechkin and I.I. Eremin who set the task to prove necessary and sufficient conditions for the existence of the committee system of linear inequalities for me in 1965.

The results of fundamental studies by RAS academician I.I. Eremin in the field of theory and methods of solutions and optimal correction of inconsistent systems of equations and inequalities, and conflicting objectives of an efficient (in particular optimal) choice determined the direction of further development of the theory and methods of operations research and pattern recognition.

One approach to this correction is associated with the collective search for generalized solutions to infeasible systems of constraints and this approach relies on different voting logic (democracy), the simplest of which is associated with decision making by majority vote [6].

The original sources of the committee theory can be found in some American works on artificial neural networks – in Nils Nilsson's, Ablow's and Keillor's algorithms [7]. However, they believed that the neural network is an engineering discipline, and therefore, they did not set the task of rigorous mathematical justification of the relevant algorithms for themselves.

There is a variety of conceptions on the basis of which decision rules for diagnosis and classification are made [8, 9].

The method of collective decisions has a wide range of applications in the area of pattern recognition and classification of objects and situations, where the learning algorithms known as committee machines, associative machines and boosting machines. Despite the apparent proximity of these approaches, for some reasons they have been developed independently for a long time [10, 11].

M.Yu. Khachay in his doctorate [12] noticed that it is possible to combine the committee theory with the theory of empirical risk.

The researches of Institute of Mechanics and Mathematics URAN are currently under development (Mazurov, Tyagunov, Kazantsev, Krivonogov, Sachkov, Beletskiy, Gaynanov, Matveev, Khachay) and their aim is to identify deep connections between these approaches that will contribute to further development of these approaches [13, 14]. So, for example, an original and profound book by D.N. Gayvanov was published which is based on combinatorial geometry and graph theory [15].

Methods of the synthesis of neural networks on the basis of the committee method have been developing.

On the basis of the problem of the minimal affine separating committee (the simplest piecewise linear classifier based on majority vote) a game theoretic approach to the development and justification of the approximate, particular polynomial learning algorithms for the recognition and classification of objects and situations is studied.

The problem of development of affine separating committee is a discrete generalization of the problem of a separating hyperplane in Euclidean space for the case of the partial sets, convex hulls of which intersect. If the partial sets intersect then the statement of this problem in this case is naturally immersed in a finite-dimensional space of a suitable dimension [16, 17].

One of the committee methods uses the analysis of finite and infinite systems of inequalities – linear and non-linear – they may be joint or disjoint. The committee method is connected with this approach [18].

However, on purpose the committee method is not limited to the separation of two finite sets of the same function. It possesses other features: there is no need to formulate separability hypotheses, and another feature is about the existence of the compactness axioms. But the only weakest necessary condition is supposed to be fulfilled: for training sets of different classes not to intersect. It is important that according to this minimal condition there is always a committee consisting of affine functions. G.S. Rubinstein noted the connection of the committee theory with the task of systems of different representatives of sets [14].

Note that the committee  $C$  is actually a set of factors.

Another approach involves the minimization of the empirical risk (V.N. Vapnik [19]). V.N. Vapnik has developed the theory of statistical learning problems. He has generalized Glivenko' theory [20] and he has developed the theory of uniform convergence of frequencies of events to their probabilities and he has introduced a diversity measure of function classes.

Y.I. Zhuravlev's approach [21, 22] – evaluation method is connected with mathematical principles of classification, this evaluation method covers many algorithms of recognition including heuristics. In particular, he develops the algebra of algorithms, including heuristic ones. And in this algebra he finds the optimal decision rule.

The great variety of books and articles is devoted to the factor analysis, but there is still some kind of mystery in this topic. There is even a completely informal part of the algorithm. This is the process when a factor obtains its meaning which combines the signs with their names. Since we deal with the names of the signs and factors, we use the methods of mathematical linguistics [23, 24].

The naming algorithm obtains a fairly complete formalization.

Here the comment by J.F. Lyotard becomes appropriate: in Sensus communities we come across with the way of thinking that is not purely philosophical or mathematical [25].

Now to the algorithm of calculation of a factor name according to the names of signs that are included in the corresponding taxon. We apply the factor analysis to the observation table of object\sign.

The first stage is creating taxon columns of signs according to their correspondence to objects. In the matrix object\sign: lines are sign values when objects are presented on the line, columns are signs when they appear in objects.

The method consists of the following: it is necessary to divide a finite set  $P$  into taxons in the space  $R^m$ . And let the form of the taxon be formulated.

The second stage for each taxon (which corresponds with the taxon object) is to write the word from the names of the signs. This word will be the name of the factor.

The third stage is the compression of a bigger word in order to convert it into the name of the factor.

Now we put it down using symbols. The observation matrix  $A$  is presented in two ways – through lines and through columns:

$$A = [C_1 \dots C_m]^* = [P_1 \dots P_n].$$

Here  $C_j^*$  – lines,  $P_i$  – columns,  $*$  – sign of transposition,  $a(c_j^*)$  – name of the object,  $a(P_i)$  – name of the sign.

Let us take one taxon  $T$  from the columns set:

$$T = \{P_i : i \in I\}.$$

The method of its finding consists of the following.  $P$  – finite set in the space  $R^m$ . And taxon form is presented:

$$T = \{x: f(x) < 0\} \cap P.$$

Here  $f$  is taken from a possible set of functions  $F$ .

Taxon  $T$  corresponds with the factor with the name  $[a(P_i): i \in I]$ . This “big” word consists of “small” words  $a(i)$ . This is the name of the factor. This word can be compressed, if needed.

It is possible to suggest another approach to the definition of the meaning of the factors. Namely the meaning of the sign or the factor is the set of their contexts. The meaning of the factor  $a$  is the set of its values or meanings  $V(a)$ . So, the meaning of the sign  $b$  is the set of its meanings or interpretations  $V(b)$ . If the factor is a combination of signs  $a$  then its meaning is in the intersection of the corresponding sets  $V(a)$ . This set is done using an electronic dictionary of synonyms.

We have already noted that the factor analysis began with the analysis of psychological researches. Let us give some information about the publications. Apparently the first work in this area belongs to F. Galton on tests theory. In 1901 K. Pearson published an article “On lines and planes of closest fit to sys-

## Краткие сообщения

---

tem of points in space”, where the idea of the principal axes was discussed. Then in 1904 Ch. Spearman published an article “General intelligence, objectively determined and measured” in “American journal of psychology”. G. Rorschach introduced psychological tests in 1921.

The factor analysis originated in sociology in 1940. From that moment the process of wide application of this data processing tool has begun.

**The research is supported by Russian Science Foundation grant no. 14-11-00109.**

### References

1. *Faktornyi, diskriminantnyi i klasternyi analiz* [Factor, Discriminant and Cluster Analysis]. Moscow, Finasy i statistika Publ., 1982. 215 p.
2. Markov. A.A. *Vvedenie v teoriyu kodirovaniya* [Introduction to Coding Theory]. Moscow, Nauka Publ., 1982. 192 p.
3. Zhukovskaya V.M., Muchnik I.B. *Faktornyi analiz v sotsial'no-ekonomicheskikh issledovaniyakh* [Factor Analysis in Socio-Economic Researches]. Moscow, Statistics Publ., 1976. 152 p.
4. Lyubishchev A.A. *Taksonomiya* [Taxonomy]. In the Book: Lyubishchev A.A. *Linii Demokrita i Platona v istorii kul'tury* [Democritus's and Plato's “Lines” in the History of the Culture]. St Petersburg, Aleteyya Publ., 2000. 256 p.
5. Mazurov V.D. *Metod komitetov v zadachakh optimizatsii i klassifikatsii*. [Committee Method in Optimization and Classification Tasks]. Moscow, Nauka Publ., 1990. 248 p.
6. Eremin I.I., Mazurov V.D., Astaf'ev N.N.. *Nesobstvennye zadachi lineynogo i vypuklogo programmirovaniya* [Improper Problems of Linear and Convex Programming]. Moscow, Nauka Publ., 1983. 336 p.
7. Nielsen N. *Obuchayushchiesya mashiny* [Training Machines]. Moscow, Mir Publ., 1968. 176 p.
8. Chomskiy N. *Sintaksicheskie struktury* [Syntax Structures]. *New in Linguistics*. Available at: <http://www.classes.ru/grammar/149.new-in-linguistics-2/source/worddocuments/12.htm>.
9. John Passmore. *Struktura i sintaksis. V knige: Passmor D. Sovremennye filosofiy* [Structure and Syntax. In the Book: John Passmore. Recent Philosophers]. Moscow, Idea-Press Publ., 1982. 999 p.
10. Platts M. *Ways of Meaning*. Available at: <https://mitpress.mit.edu/books/ways-meaning>.
11. Nikonov O.I., Chernavin F.P., Chernavin N.P. [Creation of Rating Models with Application of Committee Structures]. *Ustoychivoe razvitie rossiyskikh regionov: ekonomicheskaya politika v usloviyakh vneshnikh i vnutrennikh shokov. Sbornik materialov XII mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Sustainable Development of the Russian Regions: Economic Policy in the Conditions of External and Internal Shocks: Collection of Materials XII of the International Scientific and Practical Conference]. Ekaterinburg, 17–18 April, 2015, pp. 847–861. (in Russ.)
12. Khachay M.Yu. *Komitetnye resheniya nesovmestnykh system ogranicheniy i metody obucheniya raspoznavaniyu. Dokt. diss.* [Committee Decisions of Incompatible Systems of Restrictions and Methods of Training in Recognition. Doctoral Dissertation]. Chelyabinsk, CC RAS, 2004. 174 p.
13. Krivonogov A.I. [Some Issues of Justification of Committee Algorithms]. *Classification and Optimization in Control Tasks*. Sverdlovsk, Education and Research Center of Academy of Sciences USSR Publ., 1981, pp. 39–51. (in Russ.)
14. *Kombinatornye svoystva vypuklykh mnozhestv i grafov* [Combinatorial Properties of Convex Sets and Graphs]. Sverdlovsk, Education and Research Publ. Center, 1983. 82 p.
15. Gaynanov D.N. *Kombinatornaya geometriya i grafy v analize nesovmestnykh system i raspoznavaniya obrazov* [Combinatorial Geometry and Graphs in the Analysis of Inconsistent Systems and Pattern Recognition]. Moscow, Nauka Publ., 2014. 174 p.
16. Astaf'ev N.N. *Lineynye neravenstva i vypuklost'* [Linear Inequalities and Convexity]. Moscow, Nauka Publ., 1982. 153 p.
17. *Lineynye neravenstva i smezhnye voprosy* [Linear Inequalities and Related Systems]. Moscow, IL Publ., 1959. 472 p.
18. Mazurov V.D. *Metod komitetov v raspoznavanii obrazov* [Committee Method in Pattern Recognition]. Sverdlovsk, Education and Research Center of Academy of Sciences USSR, 1974.
19. Vapnik V.N., Chervonenkis A.Ya. *Teoriya raspoznavaniya obrazov* [Theory of Pattern Recognition]. Moscow, Nauka Publ., 1974. 416 p.

20. Glivenko V.I. Sulla determinazione empirica di probabilita. *Giornale della Institute Italiano degli Attuari*, 1933, 4, pp. 92–99.

21. Zhuravlev Yu.I. et al. [About Mathematical Principles of Classification of Objects and Phenomena]. *Discrete Analysis. Collection of Works of Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences USSR*, Novosibirsk, 1966, vol. 7, pp. 3–15. (in Russ.)

22. Zhuravlev Yu.I. (Ed.) *Raspoznavanie. Klassifikatsiya. Prognoz. Matematicheskie metody i ikh primenenie* [Recognition. Classification. Prognosis. Matematical Methods and their Application]. Moscow, Nauka Publ., 1989. 304 p.

23. Matveev A.O. *Kompleksy sistem predstaviteley v issledovanii kombinatornykh svoystv chastichno uporyadochennykh mnozhestv i nesovmestnykh sistem lineynykh neravenstv: diss. kand. fiz.-mat. nauk* [Complexes of Systems of Representatives in the Study of Combinatorial Properties of Partially Ordered Sets and Incompatible Systems of Linear Inequalities. Candidate Dissertation]. Ekaterinburg, 1994, 119 p.

24. P. Hall. On Representatives of Subsets. *Journ. Lond. Math. Soc.*, 1935, 10, pp. 26–30.

25. Lyotard F.-J. *Soderzhanie postmoderna* [The Postmodern Condition]. St. Petersburg, Aksioma Publ., 2001. 165 p.

Received 26 March 2016

УДК 658.1-50

DOI: 10.14529/ctcr160315

## ФАКТОРНЫЙ АНАЛИЗ И ПОИСК ОБЪЕКТИВНОГО СМЫСЛА ФАКТОРОВ КАК ФУНКЦИИ СМЫСЛОВ (ИМЁН) ПРИЗНАКОВ

**Вл.Д. Мазуров**

Уральский федеральный университет им. первого Президента России Б.Н. Ельцина,  
г. Екатеринбург

Рассмотрен метод факторного анализа, включающий обозначение факторов. Если  $x$  – признаки и  $f$  – факторы, то мы ищем зависимости  $f(x)$ , не  $x(f)$ . В начале статьи приведен подробный исторический обзор источников и развитие теории факторного анализа. Далее изучены методы комитетных решений задач распознавания образов, включая дискриминантный анализ, таксономию и информативные подсистемы оценки знаков. Проанализировано объединение этих методов с искусственными нейронными сетями и факторным анализом, что позволяет находить глубокие взаимосвязи в таблице наблюдений.

Представлен пошаговый алгоритм для определения имени фактора в соответствии с именами признаков, которые включены в соответствующий таксон. Вместе с этим факторный анализ применен к таблице наблюдений объект\ признак.

В заключение по предмету исследования представлен полный ссылочный список.

*Ключевые слова:* факторы, имена, признаки, лингвистика, статистика, алгебра.

### Литература

1. Факторный, дискриминантный и кластерный анализ: пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
2. Марков А.А. Введение в теорию кодирования / А.А. Марков. – М.: Наука, 1982. – 192 с.
3. Жуковская, В.М. Факторный анализ в социально-экономических исследованиях / В.М. Жуковская, И.Б. Мучник. – М.: Статистика, 1976. – 152 с.
4. Любичев, А.А. Таксономия // Любичев, А.А. Линии Демокрита и Платона в истории культуры / А.А. Любичев. – СПб.: Алетейя, 2000. – 256 с.
5. Мазуров, Вл.Д. Метод комитетов в задачах оптимизации и классификации / Вл.Д. Мазуров. – М.: Наука, 1990. – 248 с.
6. Ерёмин, И.И. Несобственные задачи линейного и выпуклого программирования / И.И. Ерёмин, Вл.Д. Мазуров, Н.Н. Астафьев. – М.: Наука, 1983. – 336 с.

## Краткие сообщения

---

7. Нильсон, Н. *Обучающиеся машины* / Н. Нильсон. – М.: Мир, 1968. – 176 с.
8. Хомский, Н. *Синтаксические структуры* / Н. Хомский // *Новое в лингвистике*. – <http://www.classes.ru/grammar/149.new-in-linguistics-2/source/worddocuments/12.htm>.
9. Пассмор, Дж. *Структура и синтаксис* // Пассмор, Дж. *Современные философы* / Дж. Пассмор. – М.: Идея-Пресс, 1982. – 999 с.
10. Platts, M. *Ways of meaning* / M. Platts. – 1997. – 320 p. – <https://mitpress.mit.edu/books/ways-meaning>.
11. Никонов, О.И. *Построение рейтинговых моделей с применением комитетных конструкций* / О.И. Никонов, Ф.П. Чернавин, Н.П. Чернавин // *Устойчивое развитие российских регионов: экономическая политика в условиях внешних и внутренних шоков: сб. материалов XII междунар. науч.-практ. конф., г. Екатеринбург, 17–18 апреля 2015 г.* – Екатеринбург: УрФУ, 2015. – С. 847–861.
12. Хачай, М.Ю. *Комитетные решения несовместных систем ограничений и методы обучения распознаванию: дис. ... д-ра физ.-мат. наук* / М.Ю. Хачай. – Челябинск: ВЦ РАН, 2004. – 174 с.
13. Кривоногов, А.И. *Некоторые вопросы обоснования комитетных алгоритмов* / А.И. Кривоногов // *Классификация и оптимизация в задачах управления*. – Свердловск: УНЦ АН СССР, 1981. – С. 39–51.
14. *Комбинаторные свойства выпуклых множеств и графов*. – Свердловск: УНЦ, 1983. – 82 с.
15. Гайнанов, Д.Н. *Комбинаторная геометрия и графы в анализе несовместных систем и распознавания образов* / Д.Н. Гайнанов. – М.: Наука, 2014. – 174 с.
16. Н. Н. Астафьев. *Линейные неравенства и выпуклость*. – М.: Наука, 1982. – 153 с.
17. *Линейные неравенства и смежные вопросы*. – М.: ИЛ, 1959. – 472 с.
18. *Метод комитетов в распознавании образов* / ред. Вл.Д. Мазуров. – Свердловск: УНЦ АН СССР, 1974.
19. Вапник, В.Н. *Теория распознавания образов* / В.Н. Вапник, А.Я. Червоненкис. – М.: Наука, 1974. – 416 с.
20. Glivenko, V.I. *Sulla determinazione empirica di probabilita* / V.I. Glivenko // *Giornale della Institute Italiano degli Attuari*, 1933, 4, pp. 92–99.
21. *О математических принципах классификации предметов и явлений* / Ю.И. Журавлёв и др. // *Дискретный анализ: сб. тр. ИМ СО АН СССР*. – Новосибирск, 1966. – Т. 7. – С. 3–15.
22. *Распознавание. Классификация. Прогноз. Математические методы и их применение* / отв. ред. Ю.И. Журавлёв. – М.: Наука, 1989. – Вып. 2. – 304 с.
23. Матвеев, А.О. *Комплексы систем представителей в исследовании комбинаторных свойств частично упорядоченных множеств и несовместных систем линейных неравенств: дис. ... канд. физ.-мат. наук* / А.О. Матвеев. – Екатеринбург, 1994. – 119 с.
24. Hall, P. *On Representatives of Subsets* / P. Hall // *Journ. Lond. Math. Soc.* – 1935, 10. – P. 26–30.
25. Лиотар, Ф.-Ж. *Содержание постмодерна* / Ф.-Ж. Лиотар. – СПб.: Аксиома. – 2001. – 165 с.

**Мазуров Владимир Данилович**, д-р физ.-мат. наук, профессор, профессор кафедры эконометрики и статистики высшей школы экономики и менеджмента, Уральский федеральный университет им. первого Президента России, Екатеринбург; [vldmazurov@gmail.com](mailto:vldmazurov@gmail.com).

*Поступила в редакцию 26 марта 2016 г.*

---

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Mazurov, V.D. The Factor Analysis and the Search for an Objective Meaning of Factors as a Function of Meanings of (Names) Features / V.D. Mazurov // *Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника»*. – 2016. – Т. 16, № 3. – С. 137–142. DOI: 10.14529/ctcr160315

### FOR CITATION

Mazurov V.D. The Factor Analysis and the Search for an Objective Meaning of Factors as a Function of Meanings of (Names) Features. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2016, vol. 16, no. 3, pp. 137–142. DOI: 10.14529/ctcr160315